**Skype Meeting 23.05.2019**

State of art

1. Fixed a lot of bugs
2. Back to beating the base policy
3. Going through the plots

Recommendations

1. Policy to check out: Softmax, instead of greedy (those methods are hard to tune)
2. Check how is the traffic allocated
3. Take instead of occupancy, try just the av speed for the state representation
4. What happens to the vector of actions?
5. Reward redefinition (?) - maybe every minute
6. Richer state representation, let the network pick for itself what it needs
7. Exploration: the more it explores, the more it learns about the model of world – cost: you are not exploiting your knowledge
8. When exploring for long, it does get to the critical situations